J. Appl. Probab. 1–18 (2024)
*doi:*10.1017/jpr.2024.18

GEOMETRIC RANDOM INTERSECTION GRAPHS WITH GENERAL CONNECTION PROBABILITIES

MARIA DEIJFEN,* *Stockholm University*
RICCARDO MICHIELAN,** *University of Twente*

Abstract

Let $\mathcal{V}$ and $\mathcal{U}$ be the point sets of two independent homogeneous Poisson processes on $\mathbb{R}^d$. A graph $\mathcal{G}_{\mathcal{V}}$ with vertex set $\mathcal{V}$ is constructed by first connecting pairs of points (v, u) with $v \in \mathcal{V}$ and $u \in \mathcal{U}$ independently with probability $g(v-u)$, where g is a non-increasing radial function, and then connecting two points $v_1, v_2 \in \mathcal{V}$ if and only if they have a joint neighbor $u \in \mathcal{U}$. This gives rise to a random intersection graph on $\mathbb{R}^d$. Local properties of the graph, including the degree distribution, are investigated and quantified in terms of the intensities of the underlying Poisson processes and the function g. Furthermore, the percolation properties of the graph are characterized and shown to differ depending on whether g has bounded or unbounded support.

Keywords: Spatial random graphs; complex networks; AB percolation; degree distribution; percolation phase transition

2020 Mathematics Subject Classification: Primary 60K35
Secondary 05C80

1. Introduction

Random intersection graphs have been popular in network modeling to describe networks arising from bipartite structures. In general, an intersection graph is constructed by assigning each vertex a subset of some auxiliary space, and then connecting two vertices if their subsets intersect; see [21]. In the context of network modeling, the auxiliary space typically consists of an additional vertex set, so that the graph is constructed based on two disjoint vertex sets. A bipartite graph is then generated by connecting vertices and auxiliary vertices in some random way and, in the second step, connecting two vertices of the intersection graph if there exists an auxiliary vertex to which they are both connected; see, e.g., [1, 8, 15, 19, 22]. In applications, the vertex set can, for instance, consist of individuals while the auxiliary vertices represent social groups, so that two individuals are connected when they share a social group. Other examples include communication units connected via cell towers, and scientists related through joint papers. However, we will throughout refer to the auxiliary vertices as *groups*.

Most models of random intersection graphs in the literature are non-geometrical. From an empirical perspective, however, there are many examples where geographical aspects are

Received 30 June 2023; accepted 24 January 2024.
* Postal address: Stockholm University, Department of Mathematics, Matematiska institutionen 106 91 Stockholm, Sweden. Email: mia@math.su.se
** Postal address: University of Twente, Department of Electrical Engineering, Mathematics and Computer Science, Hallenweg 19 7522NH Enschede, Netherlands. Email: r.michielan@utwente.nl

© The Author(s), 2024. Published by Cambridge University Press on behalf of Applied Probability Trust. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from https://www.cambridge.org/core. IP address: 3.149.234.245, on 18 Sep 2024 at 11:26:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/jpr.2024.18

likely to play a role in the formation of a network of this type. Children, for instance, are more likely to join sports clubs close to their home, and wireless communication tends to take place through cell towers in the geographical vicinity. Moreover, a notion of proximity can be induced by other features, such as individuals sharing common interests or being part of the same age group, when these features are modeled on a proper space. We hence study a model where both vertices and groups have locations in space. Specifically, we model the two vertex sets as realizations of independent homogeneous Poisson processes on $\mathbb{R}^d$. A version of such a model has been studied under the name of AB Poisson Boolean model in the special case when a vertex connects to a group if and only if their distance is less than some constant; see [14]. Our purpose is to extend this model to more general connection probabilities, corresponding to the generalization of the renowned Poisson Boolean model to the random connection model; see [17].

The number of groups shared by two given vertices in our model turns out to be Poisson distributed with a parameter that depends on the distance between the vertices. In non-spatial versions of the random intersection graph, the probability for two given vertices to share more than one group is typically very small, so that the fraction of vertices that share two or more groups is negligible. The behavior of our model is in many cases more realistic, since multiple joint groups are not unlikely for vertices that are geographically close. The downside is that geometry induces heavy dependencies in the edge formation and makes it difficult to characterize the degree distribution explicitly. Indeed, the degree of a vertex is not simply the sum of the sizes of all groups the vertex is a member of, but overlaps between the groups need to be subtracted. However, we provide a characterization of the degree distribution and illustrate its behavior with simulations. We also study the percolation properties of the graph and, when the connection probabilities have unbounded support, we observe qualitatively different behavior compared to the AB Poisson Boolean model.

1.1. Description of the model

Let $\mathcal{V}$ and $\mathcal{U}$ denote the point sets of two independent homogeneous Poisson processes on $\mathbb{R}^d$ with intensities λ and μ, respectively. The sets $\mathcal{V}$ and $\mathcal{U}$ represent the positions of vertices and groups, respectively. We will often assume that $\mathcal{V}$ has a point at the origin. This is known as the Palm version of the process and it is well known that, for a Poisson process, this has the same distribution as the original process with an added point at the origin; see [5]. We think of the origin point as a typical point of the process.

Let $|\cdot|$ denote the Euclidean norm. The connection probabilities will be based on a function $g(x)\colon \mathbb{R}^d \to [0, 1]$, with the properties that $g(x) = g(y)$ if $|x| = |y|$ and $g(x) \le g(y)$ if $|x| \ge |y|$, that is, g is a non-increasing radial function. Our graph is constructed in two steps:

(i) Each pair of points (v, u) with $v \in \mathcal{V}$ and $u \in \mathcal{U}$ are connected independently with probability $g(v - u)$. This gives rise to a bipartite graph $\mathcal{G}_{\text{bi}}$ with vertex set $\mathcal{V} \cup \mathcal{U}$.

(ii) Two points $v_1, v_2 \in \mathcal{V}$ are then connected if and only if there exists $u \in \mathcal{U}$ such that both v_1 and v_2 are connected to u in $\mathcal{G}_{\text{bi}}$. The resulting random intersection graph is denoted by $\mathcal{G}_{\mathcal{V}}$ and has vertex set $\mathcal{V}$.

It is straightforward to see that the $\mathcal{U}$-points that the origin vertex in $\mathcal{V}$ is connected to in $\mathcal{G}_{\text{bi}}$ constitute an inhomogeneous Poisson process with intensity $\mu g(x)$; cf. [17, Proposition 1.3]. The expected number of groups that the origin is a member of is hence $\mu \int_{\mathbb{R}^d} g(x)\,\mathrm{d}x$ and, to

Downloaded from https://www.cambridge.org/core. IP address: 3.149.234.245, on 18 Sep 2024 at 11:26:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/jpr.2024.18

make the model non-trivial, we will throughout assume that $0 < \int_{\mathbb{R}^d} g(x)\,\mathrm{d}x < \infty$. This integral is the L_1-norm of g and we write $\int_{\mathbb{R}^d} g(x)\,\mathrm{d}x = ||g||$.

Note that the edges in $\mathcal{G}_{\mathcal{V}}$ are not independent, since the presence of an edge between two vertices gives information about the presence of a group at a suitable location causing the connection, and this information in turn affects the presence of other edges. This causes a substantial dependence between the connections of two vertices located close to each other, while the dependence becomes weaker for vertices far apart. The fact that the edges are not independently present is an important difference compared to the standard random connection model, where each pair of points x and y in a homogeneous Poisson process is connected independently with probability $g(x-y)$; see [17].

A special case of the random connection model is when $g(x) = \mathbf{1}_{|x|<r}$ for some $r \in \mathbb{R}$. This is the well-known Poisson Boolean model, where a ball with radius $r/2$ is placed at each Poisson point and two points are then connected if their balls intersect, i.e. if they are within distance r. The percolation properties of the model in the present paper for this particular choice of g were studied in [14]. The model was then motivated as a continuum analog of so-called AB percolation on lattices, where sites are independently assigned the mark A or B and only edges between sites with different marks are kept; see, e.g., [12, 23, 24] and references therein.

1.2. Basic properties

First, consider the number of groups that the origin vertex shares with another vertex $v \in \mathcal{V}$. As pointed out above, the groups that the origin is connected to form an inhomogeneous Poisson process with intensity $\mu g(x)$. This process can be thinned further to include only those groups to which v also belongs. The probability that v belongs to a group located at x is $g(v-x)$, and the doubly thinned process containing the groups where both the origin and v are members is hence an inhomogeneous Poisson process with intensity $\mu g(x)g(v-x)$.

Proposition 1. *The groups shared by vertices located at 0 and v form an inhomogeneous Poisson process on $\mathbb{R}^d$ with intensity $\mu g(x)g(v-x)$.*

For completeness, we include a formal proof of this in Section 2. Let $N_{0,v}$ denote the number of groups shared by 0 and v. Then $N_{0,v}$ is Poisson distributed with parameter $\mu \int_{\mathbb{R}^d} g(x)g(v-x)\,\mathrm{d}x$. We write

$$f(v) = \int_{\mathbb{R}^d} g(x)g(v-x)\,\mathrm{d}x, \tag{1}$$

i.e. f is the convolution of g with itself, and note that $f(v) \le \int_{\mathbb{R}^d} g(x)\,\mathrm{d}x < \infty$ for all v. Also note that f inherits the properties of being a non-increasing radial function from g – this follows immediately from the definition of f. Two vertices are connected in the random intersection graph if they share at least one group, which gives an expression for the connection probability by computing $\mathbb{P}(N_{0,v} > 0)$.

Corollary 1. *The probability that the vertices 0 and v in $\mathcal{G}_{\mathcal{V}}$ are connected is given by $p_{0,v} := 1 - \mathrm{e}^{-\mu f(v)}$.*

Note that $p_{0,v} = 0$ if $f(v) = 0$. This occurs if g has bounded support and $|v|$ is more than double the radius of this support. If $f(v) > 0$, on the other hand, it follows from the expression for $p_{0,v}$ that the connection probability can be made as close to 1 as we wish by increasing μ. Intuitively, if it is possible for 0 and v to connect, the probability that they actually do so can be made large by increasing the group intensity.

Downloaded from https://www.cambridge.org/core. IP address: 3.149.234.245, on 18 Sep 2024 at 11:26:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/jpr.2024.18

Corollary 2. *If $f(v) > 0$, then $p_{0,v} \to 1$ as $\mu \to \infty$.*

We now turn to the degree distribution in the random intersection graph. Let D denote the degree of the origin vertex in $\mathcal{G}_{\mathcal{V}}$, that is,

$$D = \sum_{v \in \mathcal{V}} \mathbf{1}_{\{0 \leftrightarrow v\}}, \tag{2}$$

where $0 \leftrightarrow v$ means that 0 and v are connected in $\mathcal{G}_{\mathcal{V}}$. Since the edges in $\mathcal{G}_{\mathcal{V}}$ are not independent, it is difficult to characterize the degree distribution precisely, but an upper bound is easily obtained by ignoring overlaps between the groups that the origin is a member of. Below, the variable N represents the number of groups of the origin, and $\{X_i\}_{i \geq 1}$ represent their sizes.

Proposition 2. *The degree D is stochastically dominated by $\sum_{i=1}^{N} X_i$, where $\{X_i\}_{i \geq 1}$ are independent and identically distributed (i.i.d.) Poisson variables with mean $\lambda ||g||$ and N is a Poisson variable with mean $\mu ||g||$, independent of $\{X_i\}_{i \geq 1}$.*

We give the short proof of Proposition 2 in Section 2. In Section 3, we give some examples of choices of connection functions g, and illustrate the corresponding degree distributions with the aid of simulations. Note that it follows from Proposition 2 that the degree distribution has an exponentially decaying tail. In order to obtain a power-law tail, the connection probability has to be made inhomogeneous; see Section 1.4 for further comments on this. Even though the full degree distribution is complicated to characterize, the expected degree can be computed.

Proposition 3. $\mathbb{E}[D] = \lambda \int_{\mathbb{R}^d} \left(1 - \mathrm{e}^{-\mu f(y)}\right) \mathrm{d}y$.

Proof. An edge indicator in the expression (2) for the degree of the origin is a Bernoulli random variable with success probability $p_{0,v}$. The points connected to the origin hence constitute a thinned version of a Poisson process with rate λ where a point at v is kept with probability $p_{0,v}$, i.e. an inhomogeneous Poisson process with intensity $\lambda p_{0,v}$. The expected number of points in such a process is given by $\lambda \int_{\mathbb{R}^d} p_{0,v} \, \mathrm{d}v$, as claimed. □

Since $1 + r < \mathrm{e}^r$ for $r \in \mathbb{R}$, we get the upper bound

$$\mathbb{E}[D] \leq \lambda \mu \int_{\mathbb{R}^d} f(y) \, \mathrm{d}y = \lambda \mu ||g||^2.$$

The expected degree can hence be made small by decreasing either λ or μ. It also follows from the expression for $\mathbb{E}[D]$ in Proposition 3 that the expected degree grows large linearly with the intensity λ of the vertex set. Indeed, groups are likely to be large if λ is large, and the origin has a positive probability of belonging to at least one group. When λ is fixed and μ increases, on the other hand, the expected degree grows to infinity if and only if g has unbounded support. Indeed, if vertices beyond a certain range are unreachable, the degree stays bounded, but if all vertices are potentially within reach, the expected degree becomes large when the number of groups increases, since the connection probability between any two vertices then comes close to 1; cf. Corollary 2. To see this from the expression for $\mathbb{E}[D]$, let $\ell(r)$ denote the volume of a d-dimensional ball with radius r, and define $r_s = \sup\{|v| : f(v) > s\}$. The integral in the expression for $\mathbb{E}[D]$ can then be bounded as

$$\ell(r_{1/\mu})\left(1 - \mathrm{e}^{-1}\right) \leq \int_{\mathbb{R}^d} \left(1 - \mathrm{e}^{-\mu f(v)}\right) \mathrm{d}v \leq \ell(r_0),$$

Downloaded from https://www.cambridge.org/core. IP address: 3.149.234.245, on 18 Sep 2024 at 11:26:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/jpr.2024.18

where $r_0 < \infty$ when g has bounded support, while $r_{1/\mu} \to \infty$ as $\mu \to \infty$ when g has unbounded support. The different behavior of the expected degree for large μ depending on the support of g gives rise to a different phase transition for percolation, as we will see in the next subsection.

1.3. Percolation phase transition

We now turn to the question of whether there exists an infinite component in the random intersection graph $\mathcal{G}_{\mathcal{V}}$. To this end, let C denote the number of vertices in the component of the origin vertex. A straightforward coupling argument shows that the percolation function $\theta(\lambda, \mu, g) = \mathbb{P}(C = \infty)$ is increasing in μ, and we define

$$\mu_c = \mu_c(\lambda, g) = \sup\{\mu : \theta(\lambda, \mu, g) = 0\}. \tag{3}$$

A graph is said to percolate if it contains an infinite component, and it follows from ergodicity that $\mathcal{G}_{\mathcal{V}}$ percolates with probability 0 or 1 whenever $\mu < \mu_c$ or $\mu > \mu_c$; see, e.g., [17, Section 2.1].

We will fix λ and investigate $\mu_c(\lambda, g)$. Another option would be to fix μ and investigate $\lambda_c(\mu, g)$ defined in an analogous way. This would, however, give rise to a qualitatively similar picture. Specifically, we have that $\mu_c(a, g) = \lambda_c(a, g)$ for all $a > 0$. To see this, let $\mathcal{G}_{\mathcal{U}}$ be the graph obtained from $\mathcal{G}_{\text{bi}}$ in a similar way to $\mathcal{G}_{\mathcal{V}}$, but projecting on the point set $\mathcal{U}$ instead of $\mathcal{V}$. That is, the sets $\mathcal{V}$ and $\mathcal{U}$ switch roles in part (ii) of the construction of the graph, so that two groups are connected if and only if there is a vertex that is a member of both of them. We have the following equivalence result.

Lemma 1. *If* $||g|| < \infty$, *then* $\{\mathcal{G}_{\mathcal{V}}$ percolates$\} \Leftrightarrow \{\mathcal{G}_{\text{bi}}$ percolates$\} \Leftrightarrow \{\mathcal{G}_{\mathcal{U}}$ percolates$\}$.

Proof. Just note that, when $||g|| < \infty$, each vertex (group) is connected to an almost surely finite number of groups (vertices) in $\mathcal{G}_{\text{bi}}$. Hence, percolation in $\mathcal{G}_{\text{bi}}$ is equivalent to the existence of an infinite path where groups and vertices alternate. The existence of such a path implies percolation in both $\mathcal{G}_{\mathcal{V}}$ and $\mathcal{G}_{\mathcal{U}}$. On the other hand, an infinite component in $\mathcal{G}_{\mathcal{V}}$ (or $\mathcal{G}_{\mathcal{U}}$) implies percolation in $\mathcal{G}_{\text{bi}}$, since all vertices in a given component of $\mathcal{G}_{\mathcal{V}}$ (or $\mathcal{G}_{\mathcal{U}}$) belong to the same component in $\mathcal{G}_{\text{bi}}$. □

Note that $\mathcal{G}_{\mathcal{U}}$ is equivalent to a graph where the roles of λ and μ are interchanged, so that μ is the intensity of the vertex set and λ of the auxiliary vertex set. Hence it follows from the above lemma that, if there is percolation in $\mathcal{G}_{\mathcal{V}}$ for $\lambda = a$ and $\mu = b$, then there is also percolation for $\lambda = b$ and $\mu = a$. It is thus enough to fix λ and vary μ.

The percolation properties of the standard random connection model (with one single vertex set with intensity λ) are described in [17]. It is shown that, for $d \geq 2$, there is a non-trivial critical value $\tilde{\lambda}_c(g) \in (0, \infty)$ such that the graph percolates if $\lambda > \tilde{\lambda}_c(g)$ while it does not percolate if $\lambda < \tilde{\lambda}_c(g)$. Write $\tilde{\lambda}_c(r)$ for the critical value in the standard Poisson Boolean model, that is, when $g(x) = \mathbf{1}_{|x|<r}$ for some $r > 0$. In [14], the same model as in the present paper is analyzed in the Poisson Boolean setting. Let $\mu_c(\lambda, r)$ denote the critical value for this model, as defined in (3). It is shown that $\mu_c(\lambda, r) = \infty$ when $\lambda < \tilde{\lambda}_c(2r)$, while $\mu_c(\lambda, r) \in (0, \infty)$ when λ is sufficiently large, where $\lambda > \tilde{\lambda}_c(2r)$ suffices for $d = 2$.

Our main result states that, when g has unbounded support, there is a non-trivial phase transition in μ in the random intersection graph $\mathcal{G}_{\mathcal{V}}$ for any fixed value of λ. When g has bounded support, the qualitative picture is the same as for the Poisson Boolean version. We write $s_{\max} = \sup\{|x| : g(x) > 0\}$.

Downloaded from https://www.cambridge.org/core. IP address: 3.149.234.245, on 18 Sep 2024 at 11:26:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/jpr.2024.18

Theorem 1. *Consider the graph $\mathcal{G}_{\mathcal{V}}$ with $||g|| < \infty$ in dimension $d \geq 2$.*

(i) If g has unbounded support, then $\mu_c(\lambda, g) \in (0, \infty)$ for any $\lambda > 0$.

(ii) If g has bounded support, then $\mu_c(\lambda, g) = \infty$ for $\lambda < \tilde{\lambda}_c(2s_{\max})$, while $\mu_c(\lambda, g) \in (0, \infty)$ if λ is sufficiently large, where $\lambda > \tilde{\lambda}_c(2s_{\max})$ suffices in $d = 2$.

Non-percolation for small values of μ is established by dominating the exploration of our graph with a subcritical branching process. When g has bounded support, our model is dominated by a Poisson Boolean model with $r = s_{\max}$, so in this case non-percolation is also an immediate consequence of the fact that the Poisson Boolean version does not percolate for small values of μ. Percolation for large values of μ when g has unbounded support is established by introducing a 1-dependent percolation process, which is shown to percolate for large values of μ due to Corollary 2. The argument can be adapted to show percolation also when g has bounded support, provided λ is sufficiently large. Finally, percolation for large values of μ as soon as $\lambda > \tilde{\lambda}_c(2s_{\max})$ is obtained by a generalization of the argument in [14].

Below we give some suggestions for further work. The rest of the paper is then organized so that the proofs are collected in Section 2, and some examples and numerical illustrations are given in Section 3.

1.4. Further work

In this section we list some open problems and possible generalizations of the model.

1.4.1. *Critical λ for $d \geq 3$.* The percolation picture described in Theorem 1 leaves open whether percolation is possible for any $\lambda > \tilde{\lambda}_c(2s_{\max})$ in dimension $d \geq 3$ when g has bounded support. The proof in $d = 2$ does not extend to higher dimensions and it would be interesting to understand if the statement is still true or if there is some interval $(\tilde{\lambda}_c(2s_{\max}), \bar{\lambda})$ where percolation is not possible.

1.4.2. *Percolation in one dimension.* Theorem 1 covers only $d \geq 2$. When g has bounded support, it is not hard to see that percolation is not possible in $d = 1$. The case when g has unbounded support is more subtle. Results for the standard random connection model indicate that the decay of g is important. In particular, if $g(x - y) \sim |x - y|^{-\delta}$, then there is a non-trivial phase transition in the standard random connection model for $\delta \in (1, 2)$, while percolation is not possible for $\delta > 2$ nor if g decays faster than polynomially. This is shown for long-range percolation on $\mathbb{Z}$ in [18, 20] and follows in the Poisson setting from the more general results in [10, 11]. It could be worthwhile to analyze the present model in more detail in $d = 1$ for g with unbounded support and derive conditions for both percolation and non-percolation.

1.4.3. *Clustering.* One reason why random intersection graphs have been popular in the non-spatial setting is that they give rise to clustering in the graph, manifested in the presence of a large number of triangles; see, e.g., [1, 6, 19]. Empirical networks arising from social interaction often exhibit clustering, since two people with a joint friend often also make friends with each other. In non-spatial random intersection graphs, clustering arises as a consequence of the fact that, if two vertices both share a group with a third vertex, then there is a non-trivial probability that they all share the same group, which in that case means that they form a triangle. Spatial graphs typically also exhibit clustering without the intersection effect, since connecting vertices based on their distance in itself induces clustering. Therefore, clustering is not a motivating factor for the spatial random intersection graph as much as in the non-spatial case.

Downloaded from https://www.cambridge.org/core. IP address: 3.149.234.245, on 18 Sep 2024 at 11:26:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/jpr.2024.18

However, it could still be interesting to analyze the model in the present paper in this respect and compare, for instance, to the standard random connection model.

1.4.4. *Inhomogeneous versions.* Empirical networks often exhibit large elements of inhomogeneity manifested, for instance, in power-law degree distributions. In order to achieve this in the present model, the edge probabilities would have to be made more variable, for instance by assigning random weights to the vertices and letting the edge probabilities be determined by a combination of the weights and the distance between the vertices. Spatial graph models of this type without the intersection effect have been extensively studied the last few years; see, e.g., [4, 7, 9, 11, 13, 25]. Inhomogeneous versions of non-spatial random intersection graphs have been studied in [2, 3, 6]. The present model could also be extended to such a setting, and we could then analyze the effect of the weights on the degree distribution and percolation properties.

2. Proofs

We first confirm that the number of groups shared by two vertices at 0 and v constitute an inhomogeneous Poisson process, as claimed in Proposition 1.

Proof of Proposition 1. Fix a bounded Borel set $B \subset \mathbb{R}^d$, and let $N_{0,v}(B)$ denote the number of groups in B shared by 0 and v. Clearly, $N_{0,v}(B)$ and $N_{0,v}(B')$ are independent for disjoint sets B and B'. We have to prove that $N_{0,v}(B)$ is Poisson distributed with parameter $\mu \int_B g(x)g(v-x)\,\mathrm{d}x$. Write $U(B)$ for the number of groups in B, and note that $U(B)$ is Poisson distributed with parameter $\mu \cdot \ell(B)$, where $\ell(\,\cdot\,)$ denotes the Lebesgue measure on $\mathbb{R}^d$. The groups are uniformly distributed over B, and the probability that both 0 and v are members of a given group is hence given by

$$p_{0,v}(B) = \frac{\int_B g(x)g(v-x)\,\mathrm{d}x}{\ell(B)}.$$

We then have that $N_{0,v}(B) \stackrel{\mathrm{d}}{=} \sum_{i=1}^{U(B)} Z_i$, where $\{Z_i\}$ are i.i.d. Bernoulli variables with parameter $p_{0,v}(B)$, independent of $U(B)$. The generating function of $N_{0,v}(B)$ is thus given by

$$\mathbb{E}\big[s^{N_{0,v}(B)}\big] = \mathbb{E}\big[(1+p_{0,v}(B)(s-1))^{U(B)}\big] = \exp\left\{(s-1)\mu \int_B g(x)g(v-x)\,\mathrm{d}x\right\},$$

which we recognize as the generating function of the stated Poisson distribution. □

Next, we observe that the degree of a given vertex in $\mathcal{G}_{\mathcal{V}}$ can be dominated as described in Proposition 2.

Proof of Proposition 2. Let N denote the number of groups that the origin is connected to. Then N is Poisson distributed with parameter $\mu||g||$. Denote the groups of the origin by $u_1, \ldots, u_N$. The degree of the origin is given by the total number of vertices connected to these groups. We will dominate this set with the help of a superposition of independent inhomogeneous Poisson processes. First consider u_1 and let $\mathcal{P}_1$ be an inhomogeneous Poisson process with intensity function $\lambda g(x-u_1)$, representing the vertices connected to u_1 (apart from the origin). Then consider u_2 and let $\mathcal{P}_2$ be an inhomogeneous Poisson process, independent of $\mathcal{P}_1$ and with intensity function $\lambda(1-g(x-u_1))g(x-u_2)$, representing the vertices that are connected to u_2 but not to u_1. Similarly, for $i \geq 3$, the process $\mathcal{P}_i$ contains the points

Downloaded from https://www.cambridge.org/core. IP address: 3.149.234.245, on 18 Sep 2024 at 11:26:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/jpr.2024.18

connected to u_i but not to $u_1, \ldots, u_{i-1}$. For any $i \geq 1$, the intensity function of $\mathcal{P}_i$ is dominated by $\lambda g(x - u_i)$, and hence each process $\mathcal{P}_i$ can be coupled to an inhomogeneous Poisson process $\bar{\mathcal{P}}_i$ with rate $\lambda g(x - u_i)$ in such a way that the point set of the former is a subset of the latter. Write X_i for the total number of points in $\bar{\mathcal{P}}_i$. Then X_i is Poisson distributed with parameter $\lambda \int_{\mathbb{R}^d} g(x - u_i)\, \mathrm{d}x = \lambda ||g||$, and the proposition is established. □

We now turn to the percolation properties of our graph, and split the proof of Theorem 1 into a few lemmas, the first one stating that the graph does not percolate for small values of μ.

Lemma 2. *If* $||g|| < \infty$ *and* $d \geq 1$, *then* $\mu_c(\lambda, g) \in (0, \infty]$ *for any* $\lambda > 0$.

Proof. Fix $\lambda > 0$. We will explore the graph in generations, starting from the origin, and show that this exploration process is almost surely finite when μ is small. Specifically, let G_n denote the number of vertices at graph distance n from the origin. We claim that the process $\{G_n\}_{n \geq 1}$ is dominated by a branching process with offspring distribution $\sum_{i=1}^{N} X_i$, where N is a Poisson variable with parameter $\mu ||g||$ and $\{X_i\}$ are i.i.d. copies of a Poisson random variable X with parameter $\lambda ||g||$, independent of N. Indeed, it follows from Proposition 2 that G_1 is bounded in the claimed way, since G_1 is equal to the degree of the origin.

For $n \geq 2$, we iterate the same construction as in the proof of Proposition 2 and extend it to the groups. To be more precise, let $\{v_j^{(n-1)} : j = 1, \ldots, G_{n-1}\}$ denote the vertices at graph distance $n - 1$ from the origin and write $N_j^{(n-1)}$ for the number of groups that $v_j^{(n-1)}$ is a member of but where no vertex of any previous generation is a member and, for $j \geq 2$, also none of the vertices $v_1^{(n-1)}, \ldots, v_{j-1}^{(n-1)}$ in the same generation is a member. Those groups constitute an inhomogeneous Poisson process, independent of the succeeding quantities, with an intensity function that is dominated by $\mu g\left(x - v_j^{(n-1)}\right)$, where the bound is obtained by ignoring factors stemming from the exclusion of groups that have already been visited by the process. Hence the process can be coupled to an inhomogeneous Poisson process where the total number of points $\bar{N}_j^{(n-1)}$ is Poisson distributed with parameter $\mu ||g||$ in such a way that $N_j^{(n-1)} \leq \bar{N}_j^{(n-1)}$. Note that $\bar{N}_j^{(n-1)} \stackrel{\mathrm{d}}{=} N$.

Denote the groups that vertex $v_j^{(n-1)}$ is a member of by $\{u_{j,i}^{(n-1)} : i = 1, \ldots, N_j^{(n-1)}\}$ and let $X_{j,i}^{(n-1)}$ be the number of vertices that are connected to the group $u_{j,i}^{(n-1)}$ and that have not been visited before in the process. Those vertices constitute an inhomogeneous Poisson process, independent of previous quantities, with an intensity that is bounded by $\lambda g\big(x - u_{j,i}^{(n-1)}\big)$. We can hence couple the process to an inhomogeneous Poisson process where the total number of points $\bar{X}_{j,i}^{(n-1)}$ is Poisson distributed with parameter $\mu ||g||$ in such a way that $X_{j,i}^{(n-1)} \leq \bar{X}_{j,i}^{(n-1)}$. Note that $\bar{X}_{j,i}^{(n-1)} \stackrel{\mathrm{d}}{=} X$.

The contribution of each vertex $v_j^{(n-1)}$ to the next generation is $\sum_{i=1}^{N_j^{(n-1)}} X_{j,i}^{(n-1)}$. Above, we have shown that this can be bounded above by coupling the variables to independent ones with the same distributions as N and X. It follows that the exploration of the graph can be coupled to a branching process with offspring distribution $\sum_{i=1}^{N} X_i$ in such a way that the total number of vertices in the component of the origin is bounded by the total progeny in the branching process. The offspring mean is $\lambda \mu ||g||^2$ and, if λ is fixed and $||g|| < \infty$, this can be made smaller than 1 by picking μ small. The total progeny of the branching process is then finite with probability 1, and we conclude that the component of the origin is almost surely finite. □

Downloaded from https://www.cambridge.org/core. IP address: 3.149.234.245, on 18 Sep 2024 at 11:26:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/jpr.2024.18

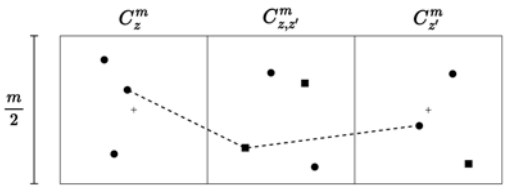


FIGURE 1. Geometrical construction in the proof of Lemma 3.

We continue by establishing percolation for large values of μ when g has unbounded support.

Lemma 3. *If g has unbounded support and $||g|| < \infty$, then $\mu_c(\lambda, g) < \infty$ for any $\lambda > 0$ and $d \geq 2$.*

The proof uses a result on domination of k-dependent bond percolation by product measure, which is an immediate consequence of the general result in [16]. Let E denote the set of nearest-neighbor edges of the $\mathbb{Z}^d$-lattice. A process $\{Y_e\}_{e\in E}$ is k-dependent if, for any two sets $E_1, E_2 \subset E$ at l_∞-distance at least k, the variables $\{Y_e\}_{e\in E_1}$ and $\{Y_e\}_{e\in E_2}$ are independent.

Lemma 4. ([16].) *For any $d \geq 2$ and $k \geq 1$, there exists $p_c = p_c(d, k) < 1$ such that, for any k-dependent process $\{Y_e\}_{e\in E}$ with $\mathbb{P}(Y_e = 1) = 1 - \mathbb{P}(Y_e = 0) > p_c$, the 1s in $\{Y_e\}_{e\in E}$ percolate almost surely.*

Proof of Lemma 3. We will define a 1-dependent bond percolation model on $\mathbb{Z}^d$ with the property that percolation in this model implies the existence of an infinite component in our graph $\mathcal{G}_{\mathcal{V}}$. The marginal probabilities in the percolation model can be made arbitrarily close to 1 by picking μ large, which will imply percolation. To define the model, for $z \in \mathbb{Z}^d$, let C_z^m denote the cube with side length $m/2$ centered at mz and, for neighboring sites z and z', let $C_{z,z'}^m$ denote the cube with side length $m/2$ centered at $m(z+z')/2$, in between C_z^m and $C_{z'}^m$; see Figure 1. Note that the maximal Euclidean distance from a point in C_z^m to a point in $C_{z,z'}^m$ is bounded from above by, for instance, dm.

Say that a site z is open if C_z^m contains at least one point of the vertex set $\mathcal{V}$. We then want to declare an edge between two sites z and z' open if both sites are open and, in addition, there is a connection between a vertex in C_z^m and a vertex in $C_{z'}^m$. This would, however, lead to long-range dependencies between edges, since we would not have control over the location of the group(s) that causes the vertices to be connected. To circumvent this, we add a restriction on the location of the connecting group. For an open cube C_z^m, let v_z denote the vertex closest to the center zm of the cube. An edge between two neighboring sites z and z' is said to be open if both sites are open and there exists a group u in the intermediate cube $C_{z,z'}^m$ such that v_z and $v_{z'}$ are both connected to u. Note that percolation of the open edges implies the existence of an infinite component in $\mathcal{G}_{\mathcal{V}}$, since open edges emanate from the same vertex v_z in a given open cube C_z^m.

The states of adjacent edges are not independent, since the edges share a vertex. However, the states of edges that do not share a vertex are defined based on Poisson configurations in disjoint regions and are therefore independent. The classification of the edges is hence 1-dependent, and it follows from Lemma 4 that the open edges percolate if the probability that an edge is open is larger than some value $p_c < 1$. We have that

$$\mathbb{P}(z \text{ open}) = \mathbb{P}\big(C_z^m \cap \mathcal{V} \neq \emptyset\big) = 1 - \mathrm{e}^{-\lambda \ell(C_z^m)}. \tag{4}$$

Downloaded from https://www.cambridge.org/core. IP address: 3.149.234.245, on 18 Sep 2024 at 11:26:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/jpr.2024.18

This probability can be made larger than $p_c^{1/3}$ by picking m large. Fix such an m. For two open sites z and z', by Proposition 1, the groups where both v_z and $v_{z'}$ are members constitute an inhomogeneous Poisson process with intensity function $\mu g(x - v_z)g(x - v_{z'})$. The number of such groups in the intermediate cube $C^m_{z,z'}$ is hence Poisson distributed with parameter

$$\mu \int_{C^m_{z,z'}} g(x - v_z)g(x - v_{z'})\,\mathrm{d}x \geq \mu\ell\big(C^m_{z,z'}\big)g(dm)^2,$$

where the bound follows since $|x - v_z| \leq dm$ and $|x - v_{z'}| \leq dm$ for $x \in C^m_{z,z'}$. Hence,

$$\mathbb{P}(\text{edge } (z, z') \text{ open} \mid z \text{ and } z' \text{ open}) \geq 1 - \mathrm{e}^{-\mu\ell\left(C^m_{z,z'}\right)g(dm)^2}. \tag{5}$$

If g has unbounded support, then $g(dm) > 0$ and thus the above probability can be made larger than $p_c^{1/3}$ by picking μ large. For m and μ chosen in this way we have

$$\mathbb{P}(\text{edge } (z, z') \text{ open}) = \mathbb{P}(z \text{ open}) \cdot \mathbb{P}(z' \text{ open}) \cdot \mathbb{P}((z, z') \text{ open} \mid z \text{ and } z' \text{ open}) > p_c. \tag{6}$$

It follows that the open edges percolate, which enforces an infinite component in $\mathcal{G}_{\mathcal{V}}$. □

Next, we observe that an analogue of Lemma 3 also holds when g has bounded support, but with the additional requirement that λ must be sufficiently large.

Lemma 5. *For any $d \geq 2$, if g has bounded support and $||g|| < \infty$, then $\mu_c(\lambda, g) < \infty$ when λ is sufficiently large.*

Proof. We adapt the proof of Lemma 3 to the case when g has bounded support. Recall that $s_{\max} = \sup\{|x| : g(x) > 0\}$. We use the same construction as in the proof of Lemma 3 but with $m = s_{\max}/(2d)$. The probability (4) that a cube C^m_z contains at least one vertex from $\mathcal{V}$ can then be made larger than $p_c^{1/3}$ by picking λ large, while the conditional edge probability (5) can be made larger than $p_c^{1/3}$ by picking μ large, since $g(s_{\max}/2) > 0$. It follows as in (6) that the probability of an edge being open is larger than p_c, and we conclude that our graph percolates almost surely. □

Finally, we adapt the proof of [14, Theorem 2.1] to show that, when $d = 2$, the model percolates for large values of μ as soon as $\lambda > \tilde{\lambda}_c(2s_{\max})$.

Lemma 6. *If $d = 2$ and g has bounded support, then $\mu_c(\lambda, g) < \infty$ for $\lambda > \tilde{\lambda}_c(2s_{\max})$.*

Proof. In the proof of [14, Theorem 2.1], the result is proved for the Poisson Boolean version of the model, where two vertices within distance $2s_{\max}$ of each other are for sure connected if there is a group in the intersection of the balls with radius $s_{\max}$ centered at the respective vertices. We need to adapt the argument to take into account that such a connection in our case exists only with a certain probability, which can be made as close to 1 as necessary by increasing the number of groups in the intersection. To this end, consider the two-dimensional lattice with vertex set $m\mathbb{Z}^2$. Let $e = (z, z')$ be an edge of the lattice and consider the rectangle R_e formed by the two squares S_z and $S_{z'}$ with side length m centered at z and z'.

Fix $\lambda > \tilde{\lambda}_c(2s_{\max})$. By [17, Theorem 3.7], the critical value $\tilde{\lambda}_c(r)$ is continuous in r. It follows that there exists $s_1 < s_{\max}$ such that $\lambda > \tilde{\lambda}_c(2s_1)$. Let $s_2 \in (s_1, s_{max})$. Also, write $\tilde{\mathcal{G}}_{\mathcal{V}}(\lambda, r)$

Downloaded from https://www.cambridge.org/core. IP address: 3.149.234.245, on 18 Sep 2024 at 11:26:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/jpr.2024.18

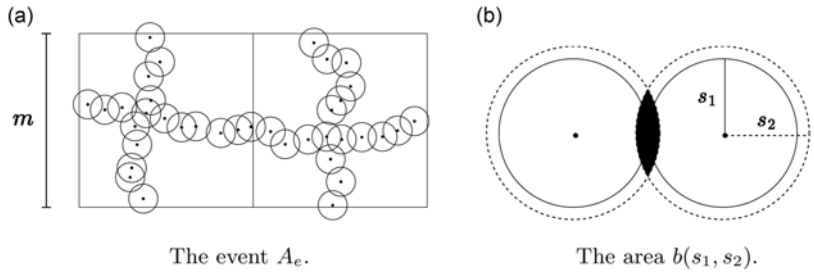


FIGURE 2. Geometrical constructions in the proof of Lemma 6.

for the graph generated by the standard Poisson Boolean model on $\mathcal{V}$. Next, we define some events associated with the edge e:

$$A_e = \{\text{there exists a crossing of } R_e \text{ along its longest side and crossings of } S_z \text{ and } S_{z'} \text{ along the perpendicular sides by a component of } \tilde{\mathcal{G}}(\lambda, 2s_1)\},$$

$$B_e = \{\text{for any two vertices } x, y \in \mathcal{V} \cap R_e \text{ such that } B_x(s_1) \cap B_y(s_1) \neq \emptyset \text{ there exists a group in } \mathcal{U} \cap B_x(s_2) \cap B_y(s_2) \text{ to which both } x \text{ and } y \text{ are connected in } \mathcal{G}_{\text{bi}}\}.$$

See Figure 2(a) for an illustration of the event A_e. The edge e is declared open if both A_e and B_e occur. The rest of the proof is divided into two steps. First we prove that the model just defined percolates provided that μ and m are large enough, and then we show that this implies percolation in $\mathcal{G}_{\mathcal{V}}$.

Step 1. Observe that the edges in the lattice form a 2-dependent process. Therefore, from Lemma 4 we know that percolation in the lattice is achieved whenever $\mathbb{P}(A_e \cap B_e) > p_{\text{c}}$. To prove the latter, we introduce an auxiliary event associated to the edge e: $V_e = \{$the number of vertices $|\mathcal{V} \cap R_e|$ is smaller than $k\}$. Trivially,

$$\mathbb{P}(A_e \cap B_e) > \mathbb{P}(A_e \cap B_e \cap V_e) = \mathbb{P}(B_e \mid A_e \cap V_e) \cdot \mathbb{P}(A_e \cap V_e).$$

Observe that $\mathbb{P}(A_e \cap V_e) > \mathbb{P}(A_e) + \mathbb{P}(V_e) - 1$. From [17, Corollary 4.1] we know that for any $\varepsilon > 0$ there exists $m = m(\varepsilon)$ such that $\mathbb{P}(A_e) > 1 - \varepsilon$, since $\lambda > \tilde{\lambda}_{\text{c}}(2s_1)$ and the graph $\tilde{\mathcal{G}}(\lambda, 2s_1)$ percolates. Moreover, the number of vertices in the rectangle $|\mathcal{V} \cap R_e|$ is Poisson distributed with mean $\lambda \cdot \ell(R_e) = 2\lambda m^2$. Thus, there exists $k = k(\varepsilon)$ such that $\mathbb{P}(V_e) > 1 - \varepsilon$. Therefore, $\mathbb{P}(A_e \cap V_e) > 1 - 2\varepsilon$ for any choice of ε, provided that m and k are large enough.

Next, we bound $\mathbb{P}(B_e \mid A_e \cap V_e)$. Given the event $A_e \cap V_e$, we know that inside R_e there are at most $k^2/2$ pairwise intersections of the balls $\{B_x(s_1)\}_{x \in \mathcal{V} \cap R_e}$. Furthermore, if the balls $B_x(s_1)$ and $B_y(s_1)$ intersect, the area $|B_x(s_2) \cap B_y(s_2)|$ can be bounded from below by $b(s_1, s_2)$, where $b(s_1, s_2)$ is the area of the lens of intersection of two balls with radius s_2 whose centers are at distance $2s_1$; see Figure 2(b). Also, observe that, for a group located in $B_x(s_2) \cap B_y(s_2)$, the probability that there is a connection to both x and y (in $\mathcal{G}_{\text{bi}}$) is bounded below by $g(s_2)^2$, since the distance to both x and y is at most s_2. The number of groups in a given intersection $B_x(s_2) \cap B_y(s_2)$, with $x, y \in \mathcal{V} \cap R_e$, is hence stochastically larger than a Poisson variable with parameter $\mu g(s_2)^2 b(s_1, s_2)$, implying that

$$\mathbb{P}(B_e \mid A_e \cap V_e) \geq \left(1 - \exp\left(-\mu g(s_2)^2 b(s_1, s_2)\right)\right)^{k^2/2}.$$

Downloaded from https://www.cambridge.org/core. IP address: 3.149.234.245, on 18 Sep 2024 at 11:26:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/jpr.2024.18

It follows that $\mathbb{P}(B_e \mid A_e \cap V_e)$ can be made as close to 1 as we wish by increasing μ. Summing up, for any $\varepsilon > 0$, there exist m, μ, k such that $\mathbb{P}(A_e \cap B_e \cap V_e) > 1 - \varepsilon$.

Step 2. Fix m and μ such that the edge percolation model on $m\mathbb{Z}^2$ defined above percolates. Suppose e_1, e_2 are two adjacent edges in the infinite component, that is, suppose the events A_{e_1}, A_{e_2}, B_{e_1}, and B_{e_2} occur. Write L_{e_1} and L_{e_2} for the crossings of the respective rectangles along the longest side. Then, as a consequence of the existence of crossings (induced by balls with radius s_1) stipulated by A_{e_1} and A_{e_2}, we have that either the two crossings L_{e_1}, L_{e_2} intersect (if the edges e_1 and e_2 have different orientations) or they both intersect a perpendicular crossing in the square $R_{e_1} \cap R_{e_2}$ (if e_1 and e_2 have the same orientation). Now draw balls of radius s_2 around each vertex of the crossings L_{e_1}, L_{e_2} (and eventually the perpendicular crossing). The events B_{e_1} and B_{e_2} imply that every pairwise intersection of these balls contains a group that is connected to the vertices at the center of the two balls in the graph $\mathcal{G}_{\text{bi}}$. It follows that all vertices in the crossings L_{e_1} and L_{e_2}, as well as the perpendicular crossing, belong to the same component in $\mathcal{G}_{\mathcal{V}}$. Therefore, $\mathcal{G}_{\mathcal{V}}$ percolates whenever the edge percolation model on $m\mathbb{Z}^2$ does. □

Remark 1. Lemma 6 can also be applied when g has unbounded support to conclude that $\mu_c(\lambda, g) < \infty$ for any $\lambda > 0$ in $d = 2$ (that is, the conclusion of Lemma 3 in $d = 2$). Indeed, for a fixed $\lambda > 0$, we can take $r < \infty$ such that $\lambda > \tilde{\lambda}_c(2r)$, and Lemma 6 then shows that a model with the connection function g truncated at r gives rise to percolation, which implies percolation in the non-truncated model as well.

Theorem 1 now follows by combining the above lemmas.

Proof of Theorem 1. Part (i) follows from Lemma 2 and Lemma 3. As for part (ii), the fact that percolation is not possible when $\lambda < \tilde{\lambda}_c(2s_{\max})$ follows by recalling from Lemma 1 and its proof that, if there is an infinite component in $\mathcal{G}_{\mathcal{V}}$, then there is an infinite component in $\mathcal{G}_{\text{bi}}$ where vertices and groups alternate. This means that there is an infinite component in the Poisson Boolean model with vertex set $\mathcal{V}$ and $r = 2s_{\max}$, since two consecutive vertices (with one intermediate group) on the infinite path in $\mathcal{G}_{\text{bi}}$ must be within distance $2s_{\max}$ from each other. But such a component does not exist when $\lambda < \tilde{\lambda}_c(2s_{\max})$. The fact that $\mu_c(\lambda, g) < \infty$ for any $d \geq 2$ when λ is large follows from Lemma 5, and Lemma 6 asserts that $\lambda > \tilde{\lambda}_c(2s_{\max})$ suffices in $d = 2$. □

3. Examples and simulations

In this section we give some examples of connection functions $g(x)$ and illustrate the behavior of the model with the help of simulations. Recall that the expected number of groups that a vertex is a member of is $\mu||g||$ and that the expected group size is $\lambda||g||$. The magnitude of $||g||$ hence plays an important role in determining the sparsity/density of the graph. To provide a fair qualitative comparison of different profile functions, we will therefore sometimes normalize g so that $||g|| = 1$. Recall the definition (1) of the self-convolution f and the role of f in the connection probability and the expected degree described in Corollary 1 and Proposition 3. Note that, when $||g|| = 1$, the function g can be interpreted as the probability density function (PDF) of a random vector Y on $\mathbb{R}^d$, and f is then the PDF of the sum of two independent realizations of Y.

Downloaded from https://www.cambridge.org/core. IP address: 3.149.234.245, on 18 Sep 2024 at 11:26:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/jpr.2024.18

We discuss three examples of profile functions:

(i) *Poisson Boolean model.* We first recall the Poisson Boolean model, where $g(x) = \mathbf{1}_{\{|x|<r\}}$, $r > 0$. The corresponding random intersection graph $\mathcal{G}_{\mathcal{V}}$ is the AB Poisson Boolean model studied in [14]. Write $B_r(v)$ for the ball with radius r centered at v. For any $v \in \mathcal{V}$, we have

$$g(x)g(v-x) = \begin{cases} 1 & \text{if } x \in B_r(0) \cap B_r(v), \\ 0 & \text{otherwise.} \end{cases}$$

Hence, $f(v) = |B_r(0) \cap B_r(v)|$ and the probability that 0 and v are connected is $p_{0,v} = 1 - \exp\{-\mu |B_r(0) \cap B_r(v)|\}$, which equals 0 when $|v| > 2r$.

(ii) *Normal distribution.* An example of connection probabilities with unbounded support is provided by

$$g(x) = \frac{1}{(2\pi\sigma^2)^{d/2}} \mathrm{e}^{-|x|^2/2\sigma^2}.$$

Here, g is the PDF of a normal random vector with $||g|| = 1$, and f is hence the PDF of the sum of two independent normal random vectors with PDF g, i.e.

$$f(v) = \frac{1}{(4\pi\sigma^2)^{d/2}} \mathrm{e}^{-|v|^2/4\sigma^2}.$$

The probability that two vertices 0 and v are connected is $p_{0,v} = 1 - \mathrm{e}^{-\mu f(v)}$, and we note that $p_{0,v} \to 1$ as $\mu \to \infty$ and decays faster than exponentially to 0 when $|v| \to \infty$.

(iii) *Power-law decay.* A connection probability function with polynomially decaying tail is given by $g(x) = 1 \wedge |x|^{-d\alpha}$, $\alpha > 1$. In the (discrete version of the) standard random connection model, this is known as long-range percolation; see the references in Section 1.4. The function f is given by

$$f(v) = \int_{\mathbb{R}^d} \left(1 \wedge |x|^{-d\alpha}\right)\left(1 \wedge |v-x|^{-d\alpha}\right) \mathrm{d}x.$$

We do not have an explicit expression for f in this case, but it can easily be estimated numerically.

3.1. Visualization

Figure 3 contains visualizations of the random intersection graph $\mathcal{G}_{\mathcal{V}}$ for the above choices of connection probabilities, scaled so that the average degree in the graphs is similar. The model has been simulated on the torus $\mathbb{T}$, represented as a square where opposite sides are identified. Vertices and groups are positioned on the torus according to two independent Poisson processes with intensity $\lambda = 2$ and $\mu = 1$. For the Poisson Boolean function and the normal distribution most edges are relatively short, due to the fact that the connection probability g decays exponentially fast with the distance. For the power-law distribution with $\alpha = \frac{3}{2}$, we observe more long-range connections in the graph.

Downloaded from https://www.cambridge.org/core. IP address: 3.149.234.245, on 18 Sep 2024 at 11:26:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/jpr.2024.18

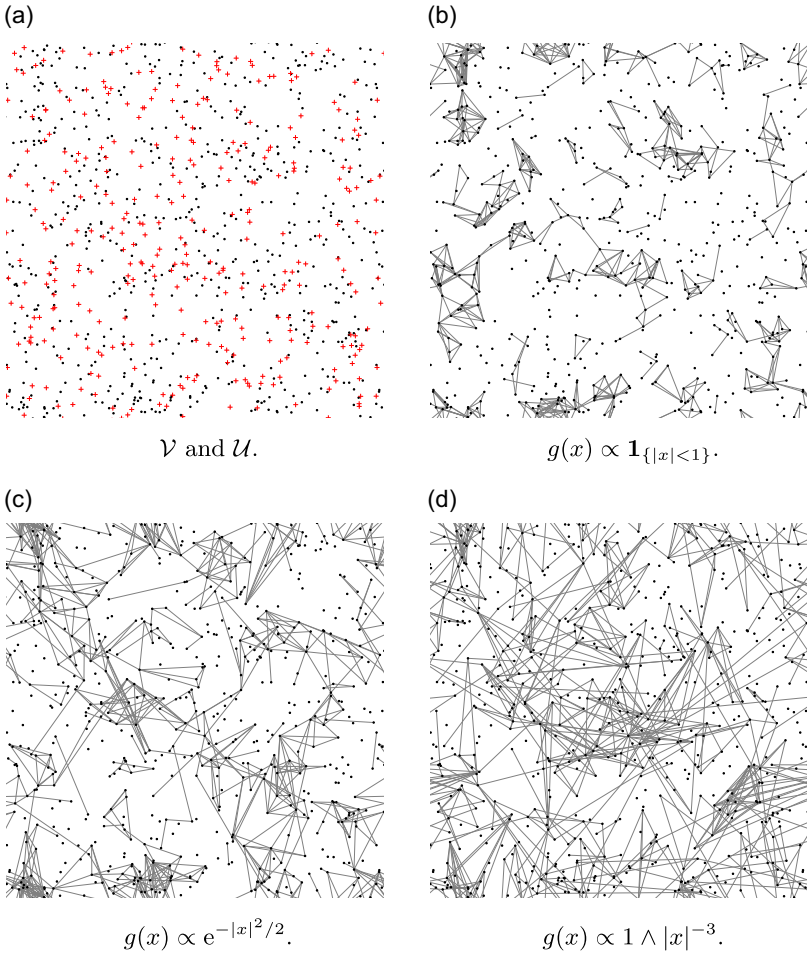
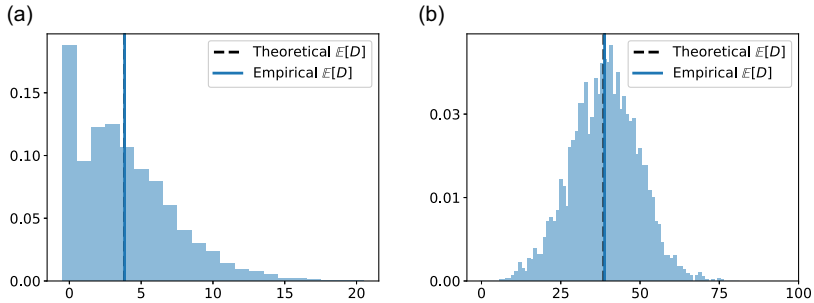


FIGURE 3. Visualization of $\mathcal{G}_{\mathcal{V}}$ for different choices of connection probabilities. The vertices and groups are sampled with intensities $\lambda = 2$, $\mu = 1$ on a torus of size 3×10^2. (a) shows the positions of vertices (black dots) and groups (red crosses). (b) (c), (d) then show the graph $\mathcal{G}_{\mathcal{V}}$ for the indicated connection probabilities.

3.2. Degree distribution

Figures 4 and 5 show the degree distribution for a few different instances of the normal distribution. Again, the model is simulated on a torus. Figure 4 illustrates the importance of $||g||$: in Figure 4(a) we have $||g|| = 1$, while $||g|| = 4$ in Figure 4(b), resulting in a degree distribution shifted towards larger degrees. In both cases, the degree distribution appears Poisson-like, apart from the spike at degree 0, most evident in Figure 4(a). This spike is explained in that a vertex that does not belong to any groups is for sure isolated, giving rise to the bound $\mathbb{P}(D = 0) \geq 1 - \mathrm{e}^{-\mu||g||}$, which is larger for smaller values of $||g||$ and μ. Figure 5 shows the degree distribution for $||g|| = 1$, with the roles of λ and μ reversed in the two pictures. Indeed, the proportion of isolated vertices is larger for the smaller value of μ in Figure 5(a).

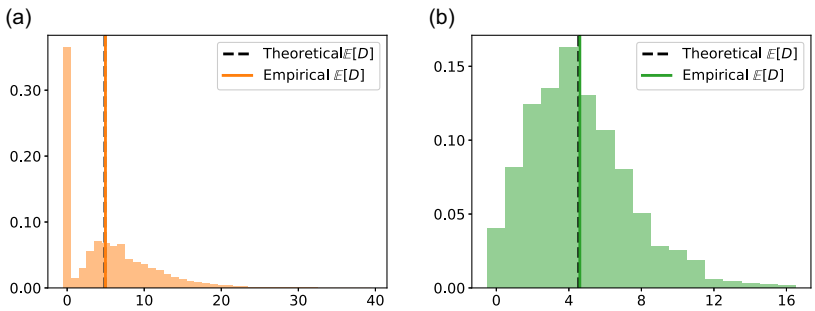
Downloaded from https://www.cambridge.org/core. IP address: 3.149.234.245, on 18 Sep 2024 at 11:26:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/jpr.2024.18



$g(x) \propto \mathrm{e}^{-|x|^2/2}$, $||g|| = 1$, $\lambda = 2$, $\mu = 2$.

$g(x) \propto \mathrm{e}^{-|x|^2/2}$, $||g|| = 4$, $\lambda = 2$, $\mu = 2$.

FIGURE 4. Degree distribution of $\mathcal{G}_{\mathcal{V}}$ sampled on a torus of size 2×10^3. The solid line is the value of the empirical expected degree, whereas the dotted line is the value of the theoretical expected degree numerically computed from the expression in Proposition 3. The intensities λ and μ are the same in (a) and (b), but the value of $||g||$ is different.



$g(x) \propto \mathrm{e}^{-|x|^2/2}$, $||g|| = 1$, $\lambda = 5$, $\mu = 1$.

$g(x) \propto \mathrm{e}^{-|x|^2/2}$, $||g|| = 1$, $\lambda = 1$, $\mu = 5$.

FIGURE 5. Degree distribution of $\mathcal{G}_{\mathcal{V}}$ sampled on a torus of size 2×10^3. The solid line is the value of the empirical expected degree, whereas the dotted line is the value of the theoretical expected degree numerically computed from the expression in Proposition 3. The intensities λ and μ are different in (a) and (b).

3.3. Percolation properties

Next, we illustrate the critical value μ_c as a function of λ and g. We simulate $\mathcal{G}_{\mathcal{V}}$ on a torus and compute the size of the largest component of $\mathcal{G}_{\mathcal{V}}$, expressed as a proportion of the total number of vertices, hoping that this provides an approximation of the component structure in infinite space. Large values of this proportion indicate percolation, whereas small values indicate no percolation. When we vary λ and μ, the critical value μ_c is approximated by the curve separating the two regions in the λ–μ plane. In Figure 6, the function g is the standard normal distribution on $\mathbb{R}^2$, and λ and μ vary between 0 and 4. We note that the two percolation phases are clearly visible and observe that $\mu_c(\lambda)$ looks symmetric with respect to the bisector of the plane, which is consistent with the exchangeability of λ and μ from Lemma 1.

Downloaded from https://www.cambridge.org/core. IP address: 3.149.234.245, on 18 Sep 2024 at 11:26:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/jpr.2024.18

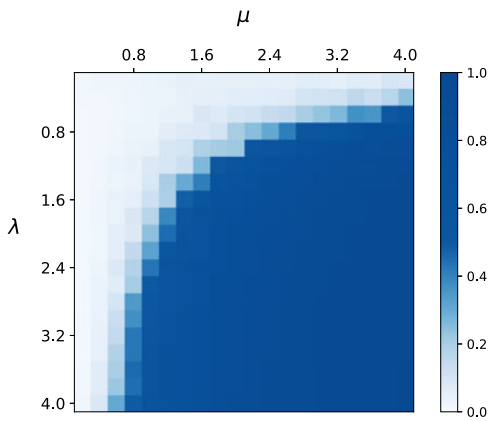


FIGURE 6. Percolation phases for $\mathcal{G}_{\mathcal{V}}$, simulated on the torus of size 10^3 using $g(x) \propto (1/2\pi)\mathrm{e}^{-|x|^2/2}$. The value in each square is the proportion of vertices in the largest component of the graph, obtained as an average over 10 samples: in the blue region the proportion of vertices in the largest component is large (percolation); in the white region the proportion is small (no percolation).

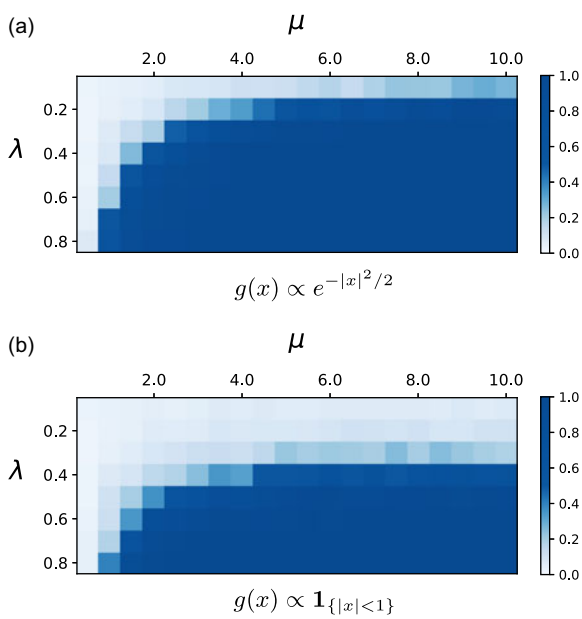


FIGURE 7. Percolation phases for $\mathcal{G}_{\mathcal{V}}$, simulated on the torus of size 10^3 averaging the proportion of vertices in the largest component over 10 samples of the graph. In (a) g has unbounded support; in (b) the support of g is bounded.

According to Theorem 1, the critical value μ_c is always finite if g has unbounded support, whereas $\mu_c = \infty$ for small values of λ if g has bounded support. This is reflected in Figure 7. Indeed, when g is the standard normal distribution on $\mathbb{R}^2$, we see that percolation also seems

Downloaded from https://www.cambridge.org/core. IP address: 3.149.234.245, on 18 Sep 2024 at 11:26:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/jpr.2024.18

achievable for small values of λ by increasing μ sufficiently, while in the Poisson Boolean, there is a value of λ below which percolation seems unfeasible.

Funding information

This work was carried out during a research visit of R. Michielan at Stockholm University in spring 2023, funded by the Erasmus+ KA131 European Mobility project. The work of M. Deijfen is supported in part by the Swedish Research Council.

Competing interests

There were no competing interests to declare which arose during the preparation or publication process of this article.

References

[1] BLOZNELIS, M. (2013). Degree and clustering coefficient in sparse random intersection graphs. *Ann. Appl. Prob.* **23**, 1254–1289.
[2] BLOZNELIS, M. (2017). Degree–degree distribution in a power-law random intersection graph with clustering. *Internet Math.* **1**.
[3] BLOZNELIS, M. AND DAMARCKAS, J. (2013). Degree distribution of an inhomogeneous random intersection graph. *Electron. J. Combinatorics* **20**, P3.
[4] BRINGMANN, K., KEUSCH, R. AND LENGLER, J. (2019). Geometric inhomogeneous random graphs. *Theoret. Comput. Sci.* **760**, 34–54.
[5] DALEY, D. AND VERE-JONES, D. (2003). *An Introduction to the Theory of Point Processes*, 2nd edn. Springer, New York.
[6] DEIJFEN, M. AND KETS, W. (2009). Random intersection graphs with tunable degree distribution and clustering. *Prob. Eng. Inf. Sci.* **23**, 661–674.
[7] DEIJFEN, M., VAN DER HOFSTAD, R. AND HOOGHIEMSTRA, G. (2013). Scale-free percolation. *Ann. Inst. H. Poincaré Prob. Statist.* **49**, 817–838.
[8] GODEHARDT, E. AND JAWORSKI, J. (2001). Two models of random intersection graphs and their applications. *Electron. Notes Discrete Math.* **10**, 129–132.
[9] GRACAR, P., HEYDENREICH, M., MÖNCH, C. AND MÖRTERS, P. (2022). Recurrence versus transience for weight-dependent random conncection model. *Electron. J. Prob.* **27**, 1–31.
[10] GRACAR, P., LÜCHTRATH, L. AND MÖNCH, C. (2022). Finitness of the percolation threshold for inhomogeneous long range models in one dimension. Preprint, arXiv:2203.11966.
[11] GRACAR, P., LÜCHTRATH, L. AND MÖRTERS, P. (2021). Percolation phase transition in weight-dependent random connection model. *Adv. Appl. Prob.* **53**, 1090–1114.
[12] HALLEY, J. (1980). AB percolation on triangular lattice. In *Ordering in Two Dimensions*, ed. S. SINHA. NORTH-HOLLAND, AMSTERDAM, pp. 369–371.
[13] HIRSCH, C. (2017). From heavy-tailed Boolean models to scale-free Gilbert graphs. *Brazilian J. Prob.* **31**, 111–143.
[14] IYER, S. AND YOGESHWARAN, D. (2012). Percolation and connectivity in AB random geometric graphs. *Adv. Appl. Prob.* **44**, 21–41.
[15] KAROŃSKI, M., SCHEINERMAN, E. AND SINGER-COHEN, K. (1999). On random intersection graphs: The subgraphs problem. *Combinatorics Prob. Comput.* **8**, 131–159.
[16] LIGGETT, T., SCHONMANN, R. AND STACEY, A. (1997). Domination by product measure. *Ann. Prob.* **25**, 71–95.
[17] MEESTER, R, and ROY, R. (1996). *Continuum Percolation*. Cambridge University Press.
[18] NEWMAN, C. AND SCHULMAN, L. (1986). One-dimensional $1/|i-j|^s$ percolation models: The existence of a transition for $s \leq 2$. *Commun. Math. Phys.* **104**, 547–571.
[19] NEWMAN, M. (2003). Properties of highly clustered networks. *Phys. Rev. E* **68**, 026121.
[20] SCHULMAN, L. (1983). Long range percolation in one dimension. *J. Phys. A* **16**, 639–641.
[21] SINGER, K. (1995). Random intersection graphs. PhD thesis, Johns Hopkins University.
[22] STARK, D. (2004). The vertex degree distribution of random intersection graphs. *Random Structures Algorithms* **24**, 249–258.

Downloaded from https://www.cambridge.org/core. IP address: 3.149.234.245, on 18 Sep 2024 at 11:26:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/jpr.2024.18

[23] WIERMAN, J. AND APPEL, M. (1987). Infinite AB percolation clusters exist on the triangular lattice. *J. Phys. A* **20**, 2533–2537.
[24] WU, X. AND POPOV, S. (2003). On AB bond percolation on the square lattice and AB site percolation on its line graph. *J. Statist. Phys.* **110**, 443–449.
[25] YUKICH, J. (2006). Ultra-small scale-free geometric networks. *J. Appl. Prob.* **43**, 665–677.

Downloaded from https://www.cambridge.org/core. IP address: 3.149.234.245, on 18 Sep 2024 at 11:26:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/jpr.2024.18